```
TGAAAGACCC CACCTGTAGG TTTGGCAAGC TAGCTTAAGT AACGCCATTT      1

TGCAAGGCAT GGAAAAATAC ATAACTGAGA ATAGAGAAGT TCAGATCAAG     51

GTCAGGAACA GATGGAACAG CTGAATATGG GCCAAACAGG ATATCTGTGG    101

TAAGCAGTTC CTGCCCCGGC TCAGGGCCAA GAACAGATGG AACAGCTGAA    151

TATGGGCCAA ACAGGATATC TGTGGTAAGC AGTTCCTGCC CCGGCTCAGG    201

GCCAAGAACA GATGGTCCCC AGATGCGGTC CAGCCCTCAG CAGTTTCTAG    251

AGAACCATCA GATGTTTCCA GGGTGCCCCA AGGACCTGAA ATGACCCTGT    301

GCCTTATTTG AACTAACCAA TCAGTTCGCT TCTCGCTTCT GTTCGCGCGC    351

TTCTGCTCCC CGAGCTCAAT AAAAGAGCCC ACAACCCCTC ACTCGGGGCG    401

CCAGTCCTCC GATTGACTGA GTCGCCCGGG TACCCGTGTA TCCAATAAAC    451

CCTCTTGCAG TTGCATCCGA CTTGTGGTCT CGCTGTTCCT TGGGAGGGTC    501

TCCTCTGAGT GATTGACTAC CCGTCAGCGG GGGTCTTTCA TTTGGGGGCT    551

CGTCCGGGAT CGGGAGACCC CTGCCCAGGG ACCACCGACC CACCACCGGG    601

AGGTAAGCTG GCCAGCAACT TATCTGTGTC TGTCCGATTG TCTAGTGTCT    651

ATGACTGATT TTATGCGCCT GCGTCGGTAC TAGTTAGCTA ACTAGCTCTG    701

TATCTGGCGG ACCCGTGGTG GAACTGACGA GTTCGGAACA CCCGGCCGCA    751

ACCCTGGGAG ACGTCCCAGG GACTTCGGGG GCCGTTTTTG TGGCCCGACC    801

TGAGTCCAAA AATCCCGATC GTTTTGGACT CTTTGGTGCA CCCCCCTTAG    851
```

FIG. 1A

AGGAGGGATA TGTGGTTCTG GTAGGAGACG AGAACCTAAA ACAGTTCCCG   901

CCTCCGTCTG AATTTTTGCT TTCGGTTTGG GACCGAAGCC GCGCCGCGCG   951

TCTTGTCTGC TGCAGCATCG TTCTGTGTTG TCTCTGTCTG ACTGTGTTTC   1001

TGTATTTGTC TGAGAATATG GGCCCGCGGG CCAGACTGTT ACCACTCCCT   1051

TAAGTTTGAC CTTAGGTCAC TGGAAAGATG TCGAGCGGAT CGCTCACAAC   1101

CAGTCGGTAG ATGTCAAGAA GAGACGTTGG GTTACCTTCT GCTCTGCAGA   1151

ATGGCCAACC TTTAACGTCG GATGGCCGCG AGACGGCACC TTTAACCGAG   1201

ACCTCATCAC CCAGGTTAAG ATCAAGGTCT TTTCACCTGG CCCGCATGGA   1251

CACCCAGACC AGGTCCCCTA CATCGTGACC TGGGAAGCCT TGGCTTTTGA   1301

CCCCCCTCCC TGGGTCAAGC CCTTTGTACA CCCTAAGCCT CCGCCTCCTC   1351

TTCCTCCATC CGCCCCGTCT CTCCCCCTTG AACCTCCTCG TTCGACCCCG   1401

CCTCGATCCT CCCTTTATCC AGCCCTCACT CCTTCTCTAG GCGCCAAACC   1451

TAAACCTCAA GTTCTTTCTG ACAGTGGGGG GCCGCTCATC GACCTACTTA   1501

CAGAAGACCC CCCGCCTTAT AGGGACCCAA GACCACCCCC TTCCGACAGG   1551

GACGGAAATG GTGGAGAAGC GACCCCTGCG GGAGAGGCAC CGGACCCCTC   1601

CCCAATGGCA TCTCGCCTAC GTGGGAGACG GGAGCCCCCT GTGGCCGACT   1651

CCACTACCTC GCAGGCATTC CCCCTCCGCG CAGGAGGAAA CGGACAGCTT   1701

CAATACTGGC CGTTCTCCTC TTCTGACCTT TACAACTGGA AAAATAATAA   1751

FIG. 1B

CCCTTCTTTT TCTGAAGATC CAGGTAAACT GACAGCTCTG ATCGAGTCTG   1801

TTCTCATCAC CCATCAGCCC ACCTGGGACG ACTGTCAGCA GCTGTTGGGG   1851

ACTCTGCTGA CCGGAGAAGA AAAACAACGG GTGCTCTTAG AGGCTAGAAA   1901

GGCGGTGCGG GGCGATGATG GGCGCCCCAC TCAACTGCCC AATGAAGTCG   1951

ATGCCGCTTT TCCCCTCGAG AATTCTACCG GGTAGGGGAG GCGCTTTTCC   2001

CAAGGCAGTC TGGAGCATGC GCTTTAGCAG CCCCGCTGGC ACTTGGCGCT   2051

ACACAAGTGG CCTCTGGCCT CGCACACATT CCACATCCAC CGGTAGCGCC   2101

AACCGGCTCC GTTCTTTGGT GGCCCCTTCG CGCCACCTTC TACTCCTCCC   2151

CTAGTCAGGA AGTTCCCCCC GCCCCGCAGC TCGCGTCGTG CAGGACGTGA   2201

CAAATGGAAG TAGCACGTCT CACTAGTCTC GTGCAGATGG ACAGCACCGC   2251

TGAGCAATGG AAGCGGGTAG GCCTTTGGGG CAGCGGCCAA TAGCAGCTTT   2301

GCTCCTTCGC TTTCTGGGCT CAGAGGCTGG GAAGGGGTGG GTCCGGGGGC   2351

GGGCTCAGGG GCGGGCTCAG GGGCGGGGCG GGCGCGAAGG TCCTCCGGAG   2401

CCCGGCATTC TGCACGCTTC AAAAGCGCAC GTCTGCCGCG CTGTTCTCCT   2451

CTTCCTCATC TCCGGGCCTT TCGACCGGAT CCGGCGATTA GTCCAATTTG   2501

TTAAAGACAG GATATCAGTG GTCCAGGCTC TAGTTTTGAC TCAACAATAT   2551

CACCAGCTGA AGCCTATAGA GTACGAGCCA TAGATAAAAT AAAAGATTTT   2601

ATTTAGTCTC CAGAAAAAGG GGGGAATGAA AGACCCCACC TGTAGGTTTG   2651

FIG. 1C

```
GCAAGCTAGC TTAAGTAACG CCATTTTGCA AGGCATGGAA AAATACATAA      2701

CTGAGAATAG AGAAGTTCAG ATCAAGGTCA GGAACAGATG GAACAGGGTC      2751

GACCCTAGAG AACCATCAGA TGTTTCCAGG GTGCCCCAAG GACCTGAAAT      2801

GACCCTGTGC CTTATTTGAA CTAACCAATC AGTTCGCTTC TCGCTTCTGT      2851

TCGCGCGCTT CTGCTCCCCG AGCTCAATAA AAGAGCCCAC AACCCCTCAC      2901

TCGGGGCGCC AGTCCTCCGA TTGACTGAGT CGCCCGGGTA CCCGTGTATC      2951

CAATAAACCC TCTTGCAGTT GCATCCGACT TGTGGTCTCG CTGTTCCTTG      3001

GGAGGGTCTC CTCTGAGTGA TTGACTACCC GTCAGCGGGG GTCTTTCATT      3051

TATGTGTCAT AAATATTTCT AATTTTAAGA TAGTATCTCC ATTGGCTTTC      3101

TACTTTTTCT TTTTATTTTT TTTTGTCCTC TGTCTCCATG TGTTGTTGTT      3151

GTTGTTTGTT TGTTTGTTTG TTGGTTGGTT GGTTAATTTT TTTTTAAAGA      3201

TCCTACACTA TAGTTCAAGC TAGACTATTA GCTACTCTGT AACCCAGGGT      3251

GACCTTGAAG TCATGGGTAG CCTGCTGTTT TAGCCTTCCC ACATCTAAGA      3301

TTACAGGTAT GAGCTATCAT TTTGGTATAT TGATTGATTG ATTGATTGAT      3351

GTGTGTGTGT GTGATTGTGT TTGTGTGTGT GATTGTGTAT ATGTGTGTAT      3401

GGTTGTGTGT GATTGTGTGT ATGTATGTTT GTGTGTGATT GTGTGTGTGT      3451

GATTGTGCAT GTGTGTGTGT GATGTGTTAG TGTATGATTG TGTGTGTGTG      3501

TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTTGT GTATATATAT      3551
```

FIG. 1D

```
TTATGGTAGT GAGAGGCAAC GCTCCGGCCC AGGCGTCAGG TTGGTTTTTG    3601

AGACAGAGTC TTTCACTTAG CTTGAATTCT TGAAGACGAA AGGGCCTCGT    3651

GATACGCCTA TTTTTATAGG TTAATGTCAT GATAATAATG GTTTCTTAGA    3701

CGTCAGGTGG CACTTTTCGG GGAAATGTGC GCGGAACCCC TATTTGTTTA    3751

TTTTTCTAAA TACATTCAAA TATGTATCCG CTCATGAGAC AATAACCCTG    3801

ATAAATGCTT CAATAATATT GAAAAAGGAA GAGTATGAGT ATTCAACATT    3851

TCCGTGTCGC CCTTATTCCC TTTTTTGCGG CATTTTGCCT TCCTGTTTTT    3901

GCTCACCCAG AAACGCTGGT GAAAGTAAAA GATGCTGAAG ATCAGTTGGG    3951

TGCACGAGTG GGTTACATCG AACTGGATCT CAACAGCGGT AAGATCCTTG    4001

AGAGTTTTCG CCCCGAAGAA CGTTTTCCAA TGATGAGCAC TTTTAAAGTT    4051

CTGCTATGTG GCGCGGTATT ATCCCGTGTT GACGCCGGGC AAGAGCAACT    4101

CGGTCGCCGC ATACACTATT CTCAGAATGA CTTGGTTGAG TACTCACCAG    4151

TCACAGAAAA GCATCTTACG GATGGCATGA CAGTAAGAGA ATTATGCAGT    4201

GCTGCCATAA CCATGAGTGA TAACACTGCG GCCAACTTAC TTCTGACAAC    4251

GATCGGAGGA CCGAAGGAGC TAACCGCTTT TTTGCACAAC ATGGGGGATC    4301

ATGTAACTCG CCTTGATCGT TGGGAACCGG AGCTGAATGA AGCCATACCA    4351

AACGACGAGC GTGACACCAC GATGCCTGCA GCAATGGCAA CAACGTTGCG    4401

CAAACTATTA ACTGGCGAAC TACTTACTCT AGCTTCCCGG CAACAATTAA    4451
```

FIG. 1E

TAGACTGGAT GGAGGCGGAT AAAGTTGCAG GACCACTTCT GCGCTCGGCC   4501

CTTCCGGCTG GCTGGTTTAT TGCTGATAAA TCTGGAGCCG GTGAGCGTGG   4551

GTCTCGCGGT ATCATTGCAG CACTGGGGCC AGATGGTAAG CCCTCCCGTA   4601

TCGTAGTTAT CTACACGACG GGGAGTCAGG CAACTATGGA TGAACGAAAT   4651

AGACAGATCG CTGAGATAGG TGCCTCACTG ATTAAGCATT GGTAACTGTC   4701

AGACCAAGTT TACTCATATA TACTTTAGAT TGATTTAAAA CTTCATTTTT   4751

AATTTAAAAG GATCTAGGTG AAGATCCTTT TTGATAATCT CATGACCAAA   4801

ATCCCTTAAC GTGAGTTTTC GTTCCACTGA GCGTCAGACC CCGTAGAAAA   4851

GATCAAAGGA TCTTCTTGAG ATCCTTTTTT TCTGCGCGTA ATCTGCTGCT   4901

TGCAAACAAA AAAACCACCG CTACCAGCGG TGGTTTGTTT GCCGGATCAA   4951

GAGCTACCAA CTCTTTTTCC GAAGGTAACT GGCTTCAGCA GAGCGCAGAT   5001

ACCAAATACT GTCCTTCTAG TGTAGCCGTA GTTAGGCCAC CACTTCAAGA   5051

ACTCTGTAGC ACCGCCTACA TACCTCGCTC TGCTAATCCT GTTACCAGTG   5101

GCTGCTGCCA GTGGCGATAA GTCGTGTCTT ACCGGGTTGG ACTCAAGACG   5151

ATAGTTACCG GATAAGGCGC AGCGGTCGGG CTGAACGGGG GGTTCGTGCA   5201

CACAGCCCAG CTTGGAGCGA ACGACCTACA CCGAACTGAG ATACCTACAG   5251

CGTGAGCTAT GAGAAAGCGC CACGCTTCCC GAAGGGAGAA AGGCGGACAG   5301

GTATCCGGTA AGCGGCAGGG TCGGAACAGG AGAGCGCACG AGGGAGCTTC   5351

FIG. 1F

CAGGGGGAAA CGCCTGGTAT CTTTATAGTC CTGTCGGGTT TCGCCACCTC 5401

TGACTTGAGC GTCGATTTTT GTGATGCTCG TCAGGGGGGC GGAGCCTATG 5451

GAAAAACGCC AGCAACGCGG CCTTTTTACG GTTCCTGGCC TTTTGCTGGC 5501

CTTTTGCTCA CATGTTCTTT CCTGCGTTAT CCCCTGATTC TGTGGATAAC 5551

CGTATTACCG CCTTTGAGTG AGCTGATACC GCTCGCCGCA GCCGAACGAC 5601

CGAGCGCAGC GAGTCAGTGA GCGAGGAAGC GGAAGAGCGC CTGATGCGGT 5651

ATTTTCTCCT TACGCATCTG TGCGGTATTT CACACCGCAT ATGGTGCACT 5701

CTCAGTACAA TCTGCTCTGA TGCCGCATAG TTAAGCCAGT ATACACTCCG 5751

CTATCGCTAC GTGACTGGGT CATGGCTGCG CCCCGACACC CGCCAACACC 5801

CGCTGACGCG CCCTGACGGG CTTGTCTGCT CCCGGCATCC GCTTACAGAC 5851

AAGCTGTGAC CGTCTCCGGG AGCTGCATGT GTCAGAGGTT TTCACCGTCA 5901

TCACCGAAAC GCGCGAGGCA GCTGCGGTAA AGCTCATCAG CGTGGTCGTG 5951

AAGCGATTCA CAGATGTCTG CCTGTTCATC CGCGTCCAGC TCGTTGAGTT 6001

TCTCCAGAAG CGTTAATGTC TGGCTTCTGA TAAAGCGGGC CATGTTAAGG 6051

GCGGTTTTTT CCTGTTTGGT CACTGATGCC TCCGTGTAAG GGGGATTTCT 6101

GTTCATGGGG GTAATGATAC CGATGAAACG AGAGAGGATG CTCACGATAC 6151

GGGTTACTGA TGATGAACAT GCCCGGTTAC TGGAACGTTG TGAGGGTAAA 6201

CAACTGGCGG TATGGATGCG GCGGGACCAG AGAAAAATCA CTCAGGGTCA 6251

FIG. 1G

```
ATGCCAGCGC TTCGTTAATA CAGATGTAGG TGTTCCACAG GGTAGCCAGC   6301

AGCATCCTGC GATGCAGATC CGGAACATAA TGGTGCAGGG CGCTGACTTC   6351

CGCGTTTCCA GACTTTACGA AACACGGAAA CCGAAGACCA TTCATGTTGT   6401

TGCTCAGGTC GCAGACGTTT TGCAGCAGCA GTCGCTTCAC GTTCGCTCGC   6451

GTATCGGTGA TTCATTCTGC TAACCAGTAA GGCAACCCCG CCAGCCTAGC   6501

CGGGTCCTCA ACGACAGGAG CACGATCATG CGCACCCGTG GCCAGGACCC   6551

AACGCTGCCC GAGATGCGCC GCGTGCGGCT GCTGGAGATG GCGGACGCGA   6601

TGGATATGTT CTGCCAAGGG TTGGTTTGCG CATTCACAGT TCTCCGCAAG   6651

AATTGATTGG CTCCAATTCT TGGAGTGGTG AATCCGTTAG CGAGGTGCCG   6701

CCGGCTTCCA TTCAGGTCGA GGTGGCCCGG CTCCATGCAC CGCGACGCAA   6751

CGCGGGGAGG CAGACAAGGT ATAGGGCGGC GCCTACAATC CATGCCAACC   6801

CGTTCCATGT GCTCGCCGAG GCGGCATAAA TCGCCGTGAC GATCAGCGGT   6851

CCAGTGATCG AAGTTAGGCT GGTAAGAGCC GCGAGCGATC CTTGAAGCTG   6901

TCCCTGATGG TCGTCATCTA CCTGCCTGGA CAGCATGGCC TGCAACGCGG   6951

GCATCCCGAT GCCGCCGGAA GCGAGAAGAA TCATAATGGG GAAGGCCATC   7001

CAGCCTCGCG TCGCGAACGC CAGCAAGACG TAGCCCAGCG CGTCGGCCGC   7051

CATGCCGGCG ATAATGGCCT GCTTCTCGCC GAAACGTTTG GTGGCGGGAC   7101

CAGTGACGAA GGCTTGAGCG AGGGCGTGCA AGATTCCGAA TACCGCAAGC   7151
```

FIG. 1H

```
GACAGGCCGA TCATCGTCGC GCTCCAGCGA AAGCGGTCCT CGCCGAAAAT    7201

GACCCAGAGC GCTGCCGGCA CCTGTCCTAC GAGTTGCATG ATAAAGAAGA    7251

CAGTCATAAG TGCGGCGACG ATAGTCATGC CCCGCGCCCA CCGGAAGGAG    7301

CTGACTGGGT TGAAGGCTCT CAAGGGCATC GGTCGACGCT CTCCCTTATG    7351

CGACTCCTGC ATTAGGAAGC AGCCCAGTAG TAGGTTGAGG CCGTTGAGCA    7401

CCGCCGCCGC AAGGAATGGT GCATGCAAGG AGATGGCGCC CAACAGTCCC    7451

CCGGCCACGG GGCCTGCCAC CATACCCACG CCGAAACAAG CGCTCATGAG    7501

CCCGAAGTGG CGAGCCCGAT CTTCCCCATC GGTGATGTCG GCGATATAGG    7551

CGCCAGCAAC CGCACCTGTG GCGCCGGTGA TGCCGGCCAC GATGCGTCCG    7601

GCGTAGAGCG CCACAGGACG GGTGTGGTCG CCATGATCGC GTAGTCGATA    7651

GTGGCTCCAA GTAGCGAAGC GAGCAGGACT GGGCGGCGGC CAAAGCGGTC    7701

GGACAGTGCT CCGAGAACGG GTGCGCATAG AAATTGCATC AACGCATATA    7751

GCGCTAGCAG CACGCCATAG TGACTGGCGA TGCTGTCGGA ATGGACGATA    7801

TCCCGCAAGA GGCCCGGCAG TACCGGCATA ACCAAGCCTA TGCCTACAGC    7851

ATCCAGGGTG ACGGTGCCGA GGATGACGAT GAGCGCATTG TTAGATTTCA    7901

TACACGGTGC CTGACTGCGT TAGCAATTTA ACTGTGATAA ACTACCGCAT    7951

TAAAGCTTTG CTTAGGAGTT TCCTAATACA TCCCAAACTC AAATATATAA    8001

GCATTTGACT TGTTCTATGC CCTAGGGGGA GGGGGGAAGC TAAGCCAGCT    8051
```

FIG. 1I

TTTTTTAACA TTTAAAATGT TAATTCCATT TTAAATGCAC AGATGTTTTT   8101

ATTTCATAAG GGTTTCAATG TGCATGAATG TCGCAATATC CTGTTACCAA   8151

AGCTAGTATA AATAAAAATA GATAAACGTG GAAATTACTT AGAGTTTCTG   8201

TCATTAACGT TTCCTTCCTC AGTTGACAAC ATAAATGCGC TGCTGAGAAG   8251

CCAGTTTGCA TCTGTCAGGA TCAATTTCCA TTATGCCAGT CATATTAATT   8301

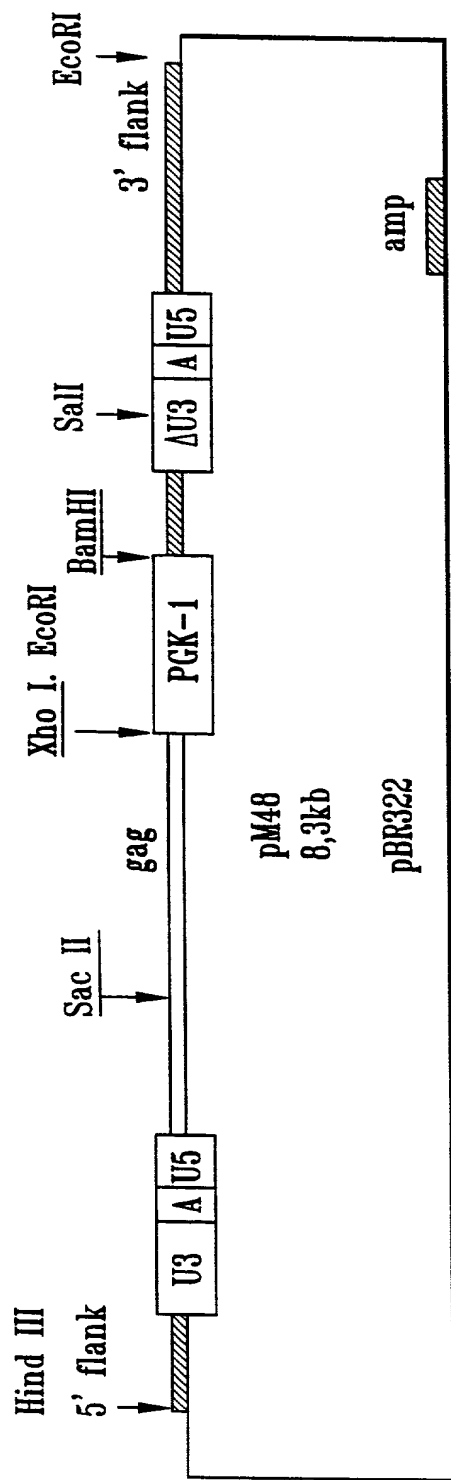ACTAGTCAAT TAGTTGATTT TTGACATATA CATGTGAA

FIG. 1J

Hind III
5' flank

U3 | A | U5

Sac II

gag

Xho I. EcoRI

PGK-1

BamHI

SalI

ΔU3 | A | U5

3' flank

EcoRI

pM48
8,3kb

pBR322

amp

FIG.2